LMA flash algorithm comparisons; clustering visualization and analysis

Eric Bruning

September, 2009

Big-picture issues

- Optimal criteria over 3D part of network
- Nonuniform detection efficiency with azimuth / range
- Variable distance/time criteria with range
- What code do we operationalize for the PG?
  - One algorithm, one parameter set
  - One algorithm, per-network parameters
  - Per-network algorithm and parameters

Collected suggestions

- Per-network tuning is important
- Kuhlman: Performance on decimated data unknown
- Carey: Define what we collectively think are desirable characteristics of good, robust flash algorithms. We could also carefully summarize characteristics of available flash rate algorithms so that they could be compared and contrasted.
- MacGorman: Outside 3D range, need to look at varying distance criteria.
    - Should be easy to test by running a variety of storms at different ranges to see where the stability point emerges

More suggestions

- Stano made several good points that are helping to persuade me that we should consider the option of having a best-for-each-network algorithm vs. a single 'winner.' The important part is to characterize the performance for each network
- Stano: most flash algorithms same for low flash rate env, but all have trouble in high rates.
    - High rate storms are the most interesting
- Stano points out that NSSTC algorithm is tuned to NALMA
    - Azimuthal and radial error characteristics, dependence on number and location of active sensors, ch 3/4 vs. ch 10
- Need to document these tunings
- Stano proposed: collect detailed error stats for each network and adapt most flexible algorithm

Work so far

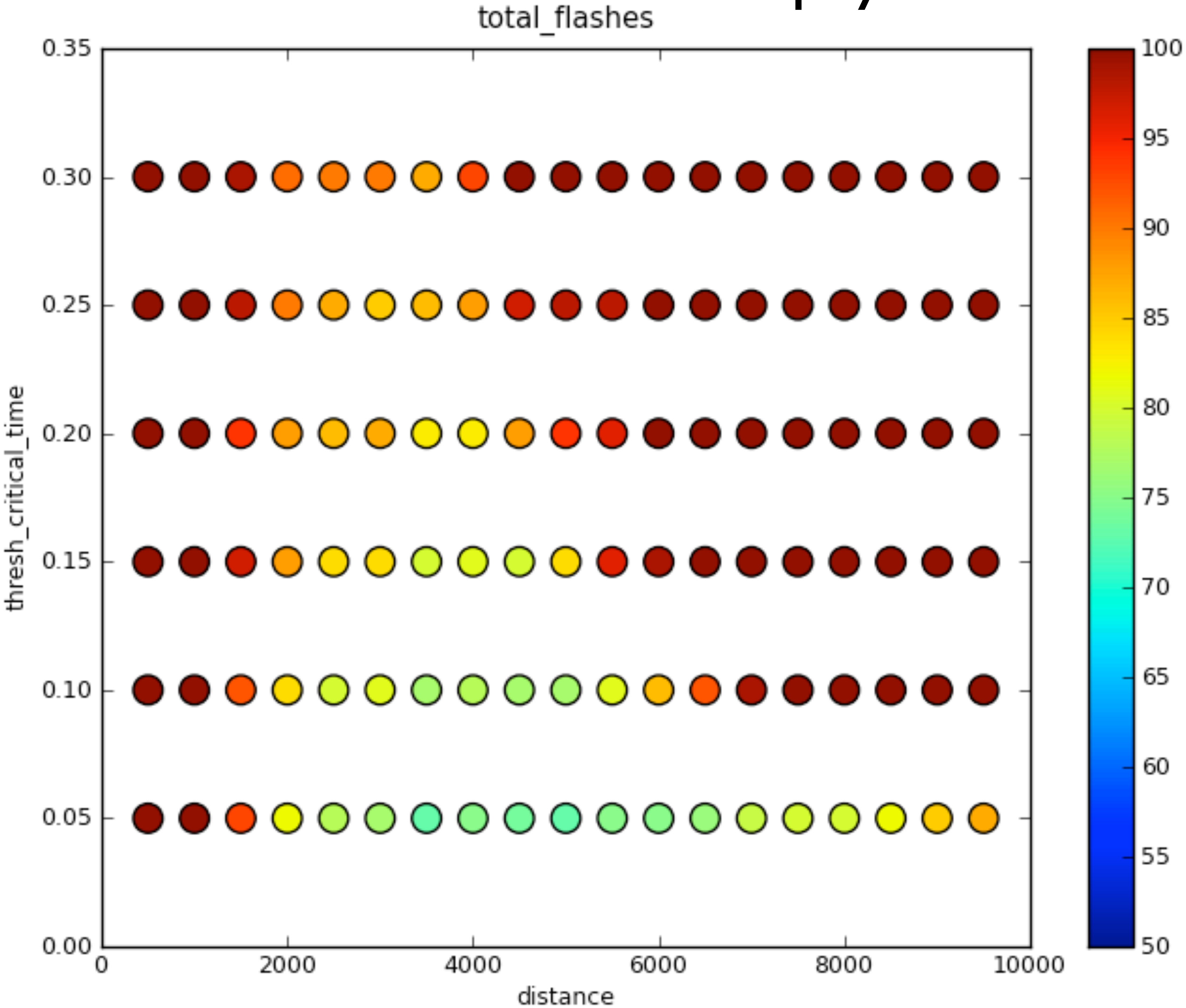
- Automation of NSSL flash algorithm
  - Variable time/space parameters
- Output rewritten to standard file format
  - Visualization / statistics code is generic
- Would like to automate testing of other algorithms, especially NSSTC. Since we're targeting ops use, probably need to port IDL algorithms to a compiled language prior to use

Clustering analysis software: Purpose and features

- Understand results of identical input to different clustering algorithm runs
- Modular code suits any clustered dataset
    - e.g., LCFA EGF, LMA flashes, pixels→cells
    - Code is available

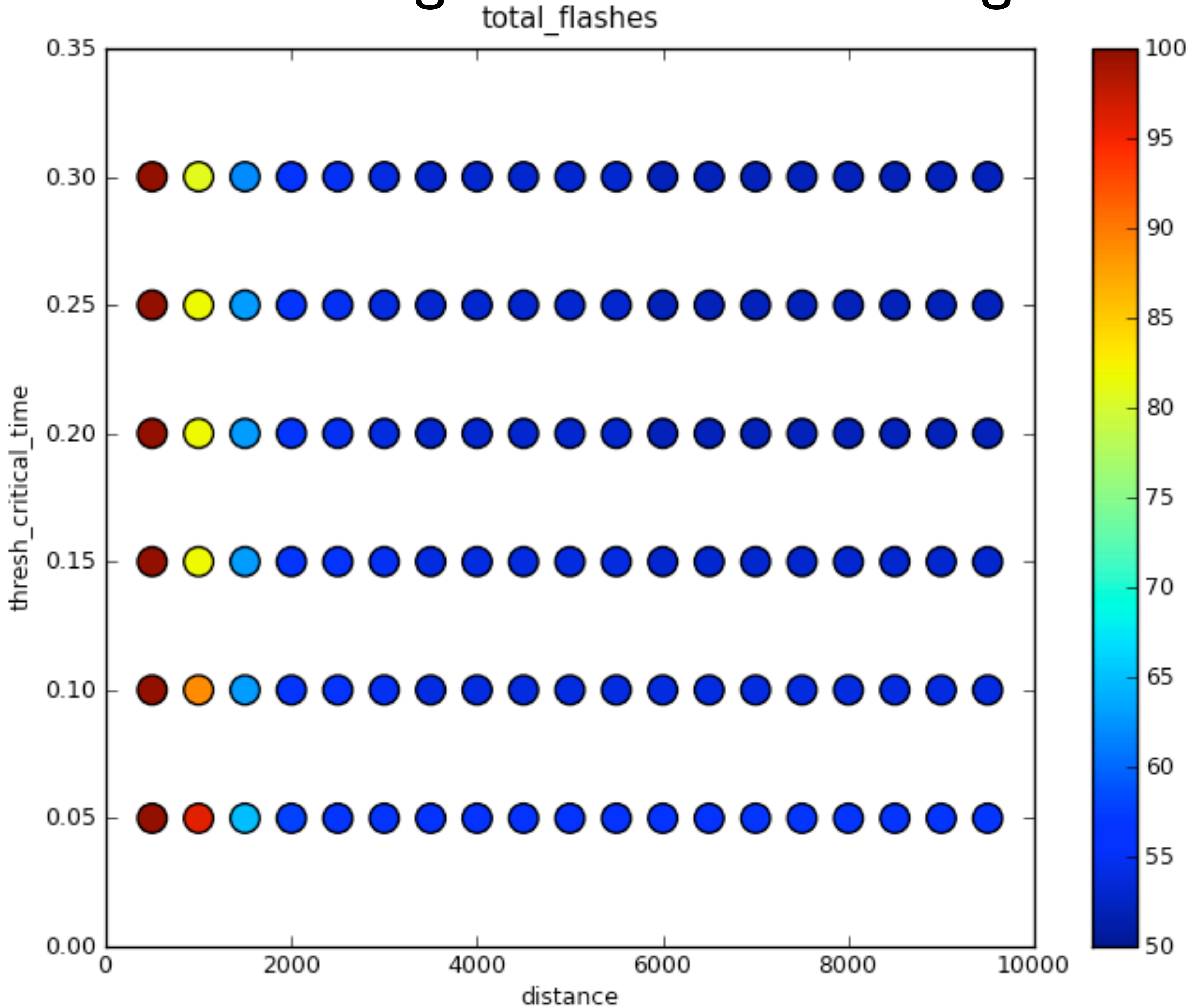
Min 6 stations, > 9 pts/fl

Patterned after Martin Murphy's 2006 study



Min 7 stations, > 9 pts/fl

Correct filtering of noise makes a big difference



Automated running

Script to the right varies parameters and produces the 114 flash algorithm runs that comprised the figures on the previous slides.

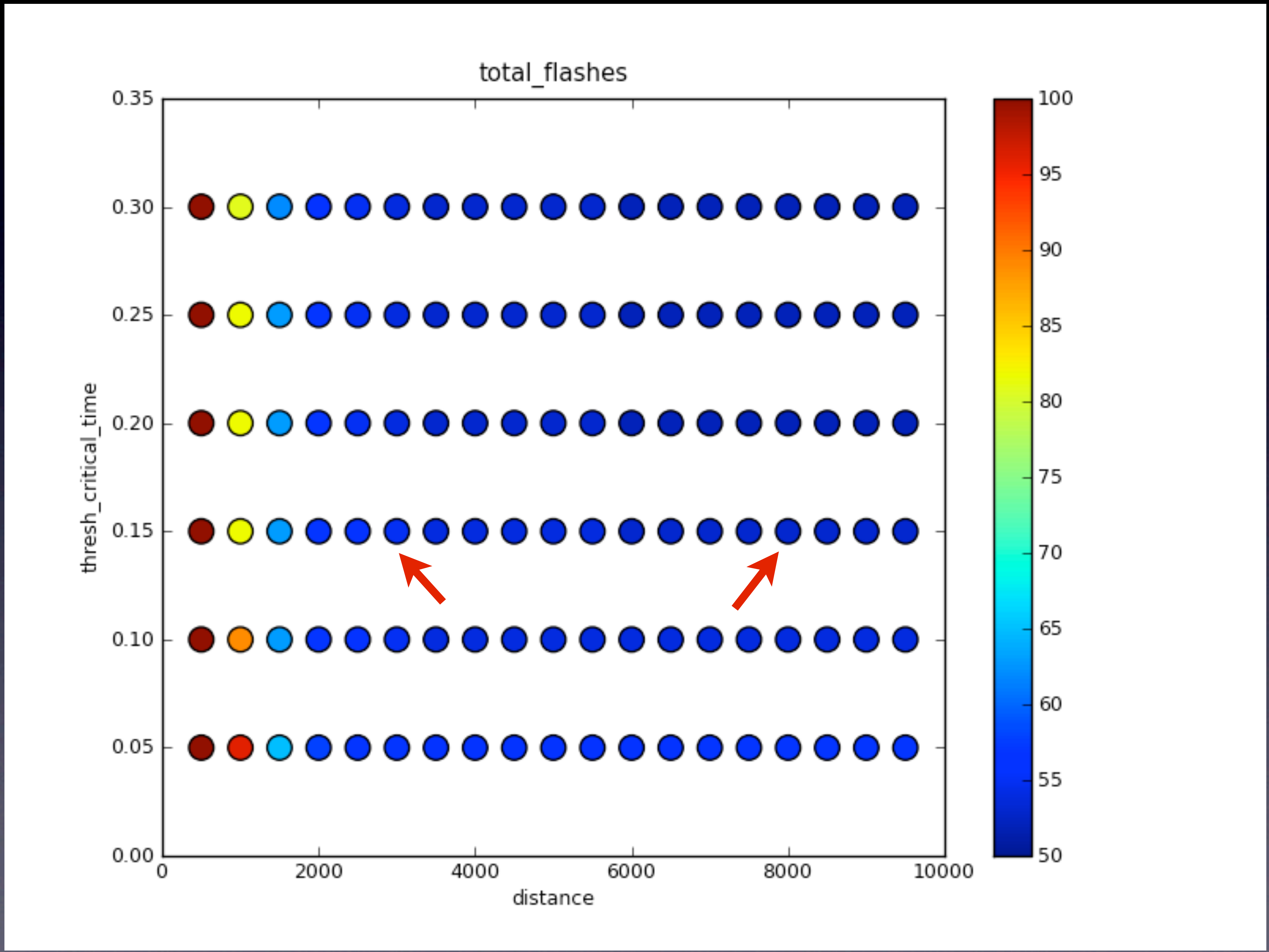
Plotting is another short canned script. Point it at the directory where the flash runs are collected.

untitled

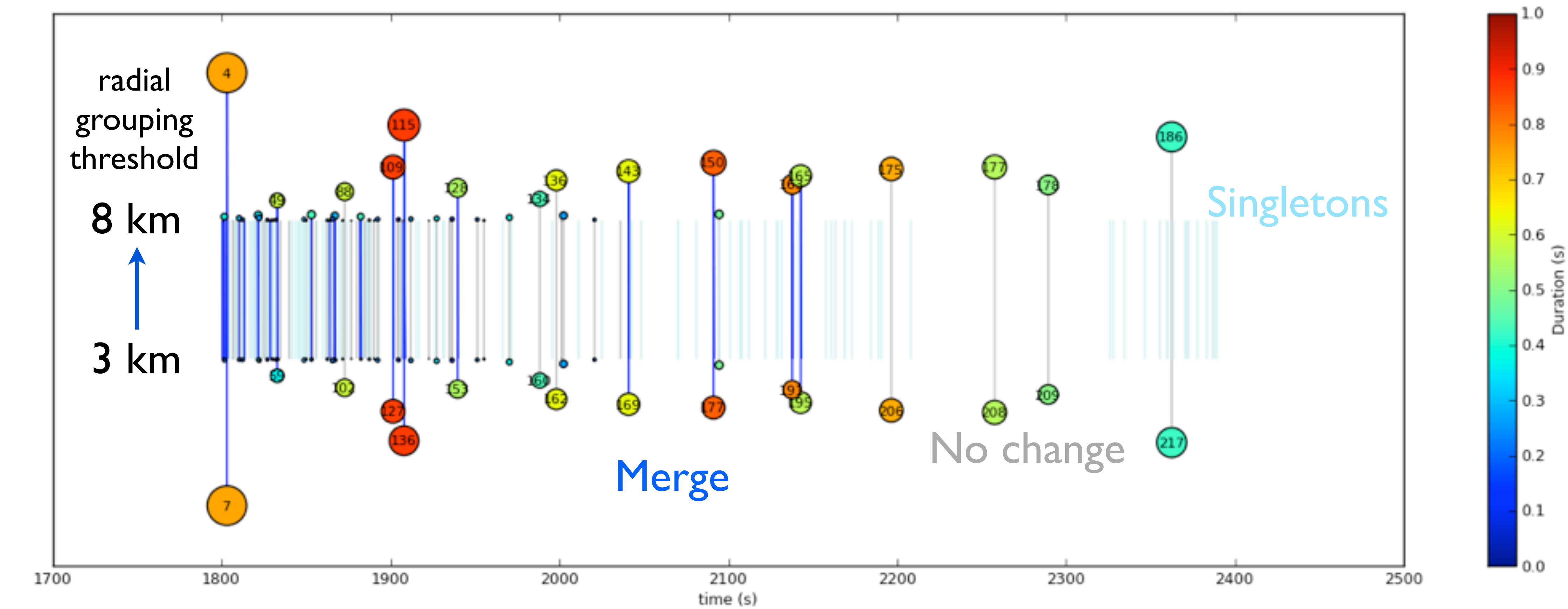
```python
import numpy as np
import os

from autorun import run_files_with_params

params = {'stations':(7,13),
          'chi2':(0,2.0),
          'lat':(-90., 90.),
          'lon':(-180., 180.),
          'alt':(0.,20000.),
          'distance':3000.0, 'thresh_duration':3.0,
          'thresh_critical_time':0.15, 'merge_critical_time':0.5,
          }

files = [
       # '/data/20040526/LMA/LYLOUT_040527_002000_0600.dat.gz',
       '/data/20040526/LMA/LYLOUT_040527_003000_0600.dat.gz',
       # '/data/20040526/LMA/LYLOUT_040527_004000_0600.dat.gz',
      ]


base_out_dir = '/Users/ebruning/out/flash_sort/'

params['lat'] = (35., 36.)

for thresh_critical_time in np.arange(0.05, 0.35, 0.05):
    for distance in np.arange(500, 10000, 500):
        params['thresh_critical_time'] = thresh_critical_time
        params['distance'] = distance

        tag = 'thresh-%s_dist-%s' % (thresh_critical_time, distance)
        outdir = os.path.join(base_out_dir, tag)
        os.mkdir(outdir)

        info = open(os.path.join(outdir, 'input_params.py'), 'w')
        info.write(str(params))
        info.close()

        run_files_with_params(files, outdir, params)
```

Line: 39 Column: 53 Python Soft Tabs: 4

Differences between two runs

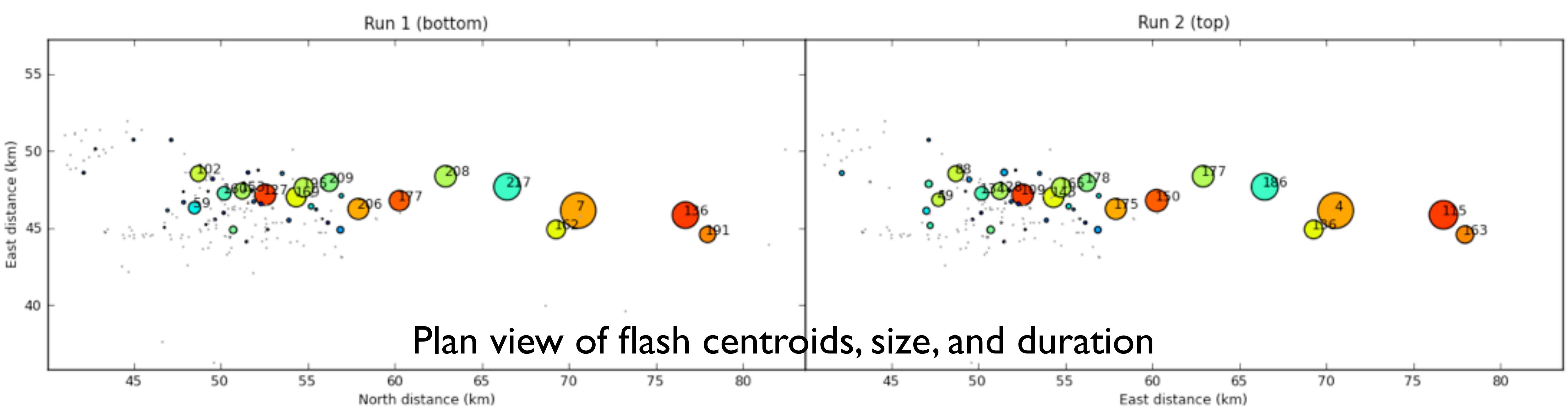


Clustering visualization



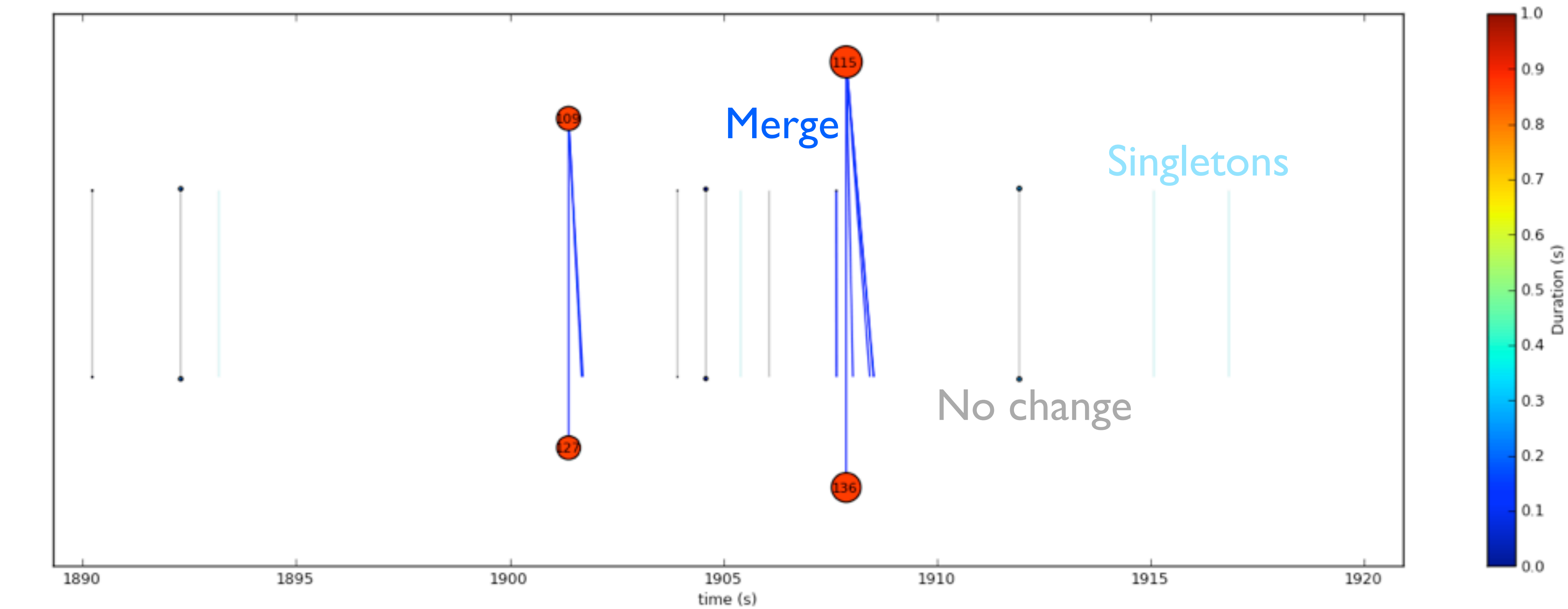
Dot size and vertical position indicate area, color indicates duration

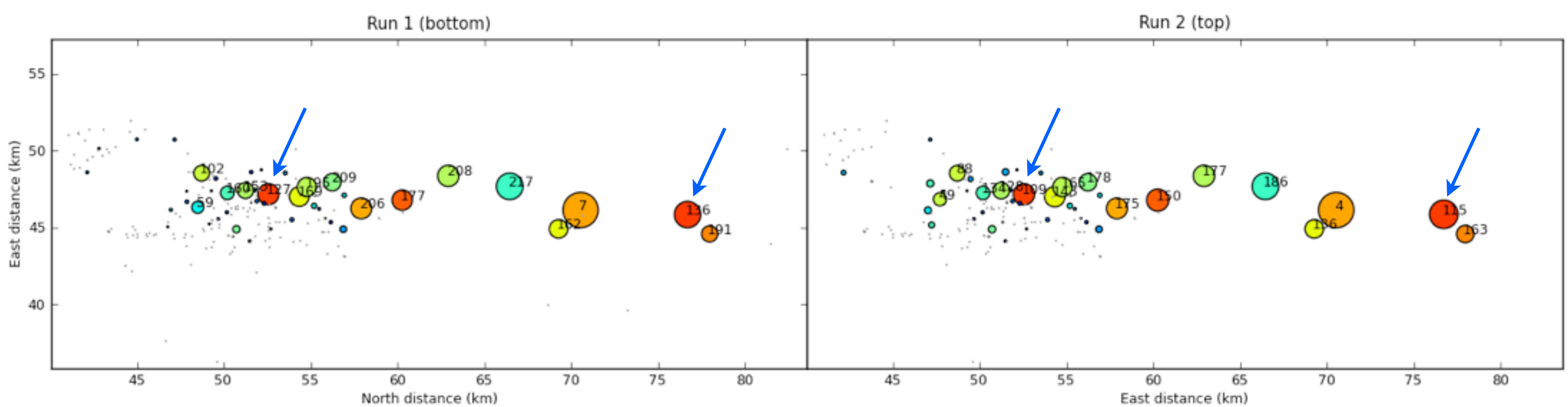
(easily configurable)



Plan view of flash centroids, size, and duration

A closer look at merging





Proving Ground / Spring Program: Work Plan and Flowchart for Operational GLM proxy data

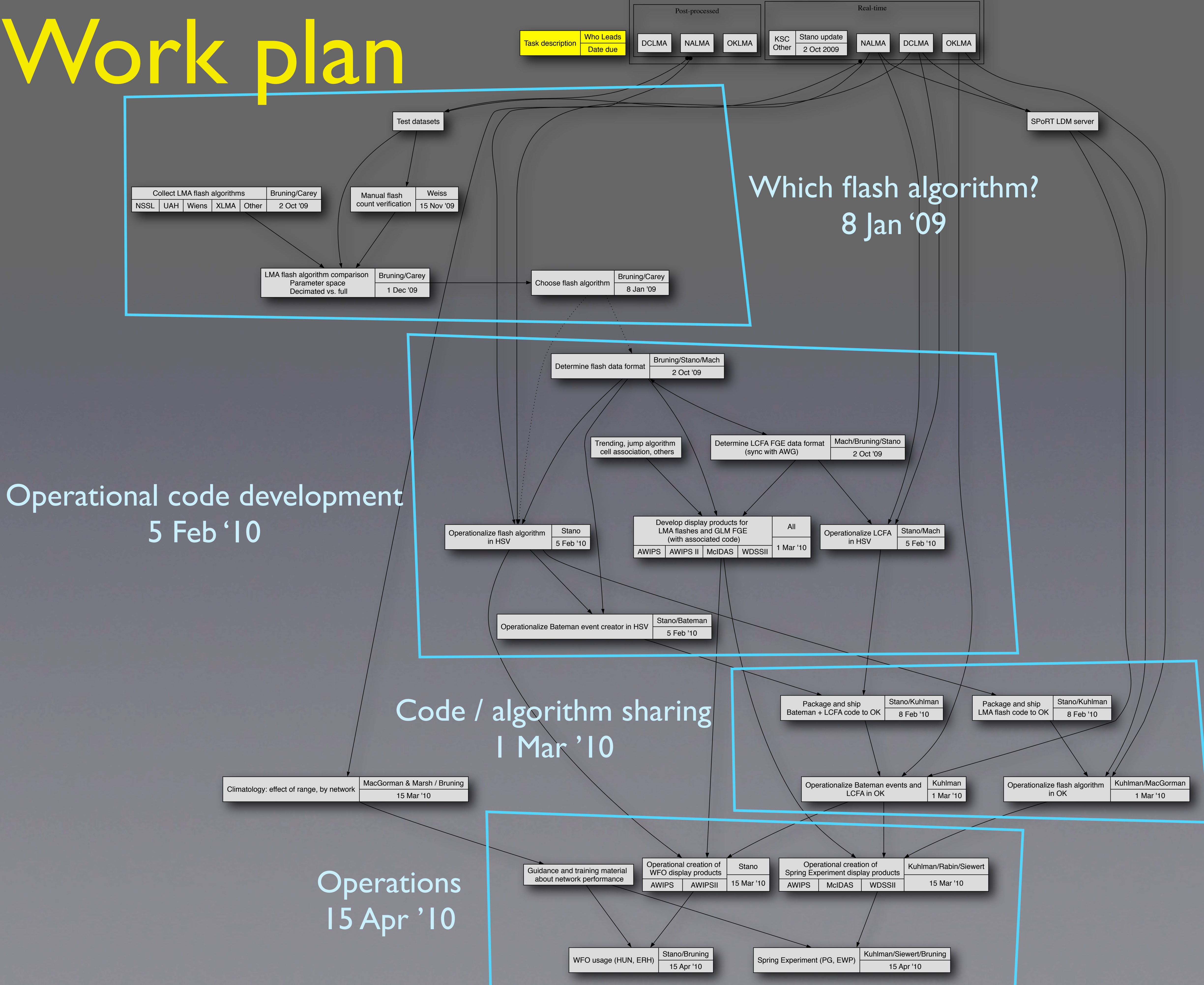
Eric Bruning

September, 2009

Conceptual pieces

- LMA sources to LMA flashes
  - Network uniformity / quality per network and optimal flash algorithm. Open documentation for end users, scientists
- LMA flashes to GLM events
  - Bateman has this underway for AL, OK. Can we work DC in too?
- LCFA: GLM events to GLM groups and flashes
- Display of LCFA and LMA flash output
  - Many target display systems

Work plan



Task description
Who Leads
Date due
Post-processed
DCLMA
NALMA
OKLMA
Real-time
KSC Other
Stano update
2 Oct 2009
NALMA
DCLMA
OKLMA
Test datasets
SPoRT LDM server
Collect LMA flash algorithms
Bruning/Carey
NSSL
UAH
Wiens
XLMA
Other
2 Oct '09
Manual flash count verification
Weiss
15 Nov '09
Which flash algorithm?
8 Jan '09
LMA flash algorithm comparison
Parameter space
Decimated vs. full
Bruning/Carey
1 Dec '09
Choose flash algorithm
Bruning/Carey
8 Jan '09
Determine flash data format
Bruning/Stano/Mach
2 Oct '09
Trending, jump algorithm cell association, others
Determine LCFA FGE data format (sync with AWG)
Mach/Bruning/Stano
2 Oct '09
Operational code development
5 Feb '10
Operationalize flash algorithm in HSV
Stano
5 Feb '10
Develop display products for LMA flashes and GLM FGE (with associated code)
All
1 Mar '10
AWIPS
AWIPS II
McIDAS
WDSSII
Operationalize LCFA in HSV
Stano/Mach
5 Feb '10
Operationalize Bateman event creator in HSV
Stano/Bateman
5 Feb '10
Package and ship Bateman + LCFA code to OK
Stano/Kuhlman
8 Feb '10
Package and ship LMA flash code to OK
Stano/Kuhlman
8 Feb '10
Code / algorithm sharing
1 Mar '10
Climatology: effect of range, by network
MacGorman & Marsh / Bruning
15 Mar '10
Operationalize Bateman events and LCFA in OK
Kuhlman
1 Mar '10
Operationalize flash algorithm in OK
Kuhlman/MacGorman
1 Mar '10
Operations
15 Apr '10
Guidance and training material about network performance
Operational creation of WFO display products
Stano
AWIPS
AWIPSII
15 Mar '10
Operational creation of Spring Experiment display products
Kuhlman/Rabin/Siewert
AWIPS
McIDAS
WDSSII
15 Mar '10
WFO usage (HUN, ERH)
Stano/Bruning
15 Apr '10
Spring Experiment (PG, EWP)
Kuhlman/Siewert/Bruning
15 Apr '10

To do

- Mark up flowchart to agree on a final schedule and any deliverables
  - Who gives what to whom?
- Common-structured file formats (NetCDF or HDF5?)
  - For both LMA flashes, GLM FGE
  - Fosters easy exchange of data and prep for ops
  - Possible inspiration from AWG framework format, AWIPS II, McIDAS-V, personal research code

Microwave precipitation estimation, lightning and the conv/strat classifciation

Wang, Bruning, Albrecht

Proposed study using TRMM

with Albrecht, Wang, Gopalan, Ferraro, others

Proposed study using TRMM

with Albrecht, Wang, Gopalan, Ferraro, others

- Improve microwave rain rates (NASA-PMM), which are also used to train of SCaMPR infrared T_b relationship (GOES-R3)

Proposed study using TRMM

with Albrecht, Wang, Gopalan, Ferraro, others

- Improve microwave rain rates (NASA-PMM), which are also used to train of SCaMPR infrared T_b relationship (GOES-R3)
- Physical basis: Passive microwave imagers primarily rely on ice-phase scattering to estimate precipitation over land. Therefore, both lightning and passive microwave are sensitive to storms with significant ice-phase microphysics.

Proposed study using TRMM

with Albrecht, Wang, Gopalan, Ferraro, others

- Improve microwave rain rates (NASA-PMM), which are also used to train of SCaMPR infrared T_b relationship (GOES-R3)
- Physical basis: Passive microwave imagers primarily rely on ice-phase scattering to estimate precipitation over land. Therefore, both lightning and passive microwave are sensitive to storms with significant ice-phase microphysics.
- An empirical P(C) to T_b relationship to for various microwave channels is determined by training against PR P(C). Retrain by also including a variety of lightning predictors.

Proposed study using TRMM

with Albrecht, Wang, Gopalan, Ferraro, others

- Improve microwave rain rates (NASA-PMM), which are also used to train of SCaMPR infrared T_b relationship (GOES-R3)
- Physical basis: Passive microwave imagers primarily rely on ice-phase scattering to estimate precipitation over land. Therefore, both lightning and passive microwave are sensitive to storms with significant ice-phase microphysics.
- An empirical P(C) to T_b relationship to for various microwave channels is determined by training against PR P(C). Retrain by also including a variety of lightning predictors.
  - e.g., flash centroid density, flash extent density, group density, flash/group density, etc.

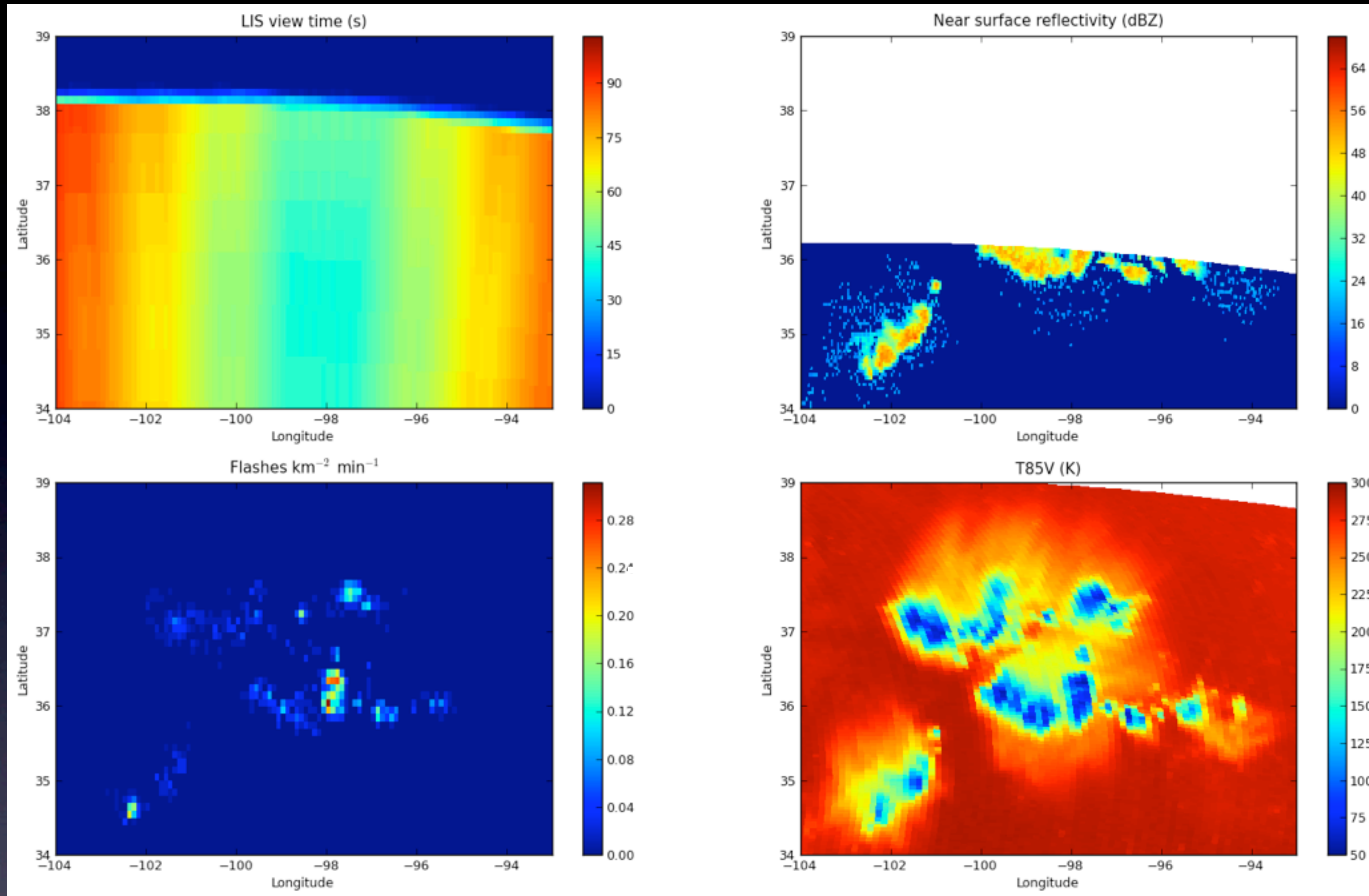
Initial work plan

- Align LIS data with TMI and PR on a common grid. Test methodology with a variety of data / cases from the U. Utah precip features database.
  - Targeted completion by October PMM meeting
- Reduce many possible parameters (flash density, groups per flash, total radiance, etc.) to a handful that PR to eliminate redundancy and emphasize subtle differences among parameters.
- On multi-year dataset, retrain TMI Convective Percentage Index using added lightning predictors (McCollum and Ferraro 2003 used stepwise linear regression)

Oklahoma, 20 June 2007

LIS view time

LIS flash centroid density



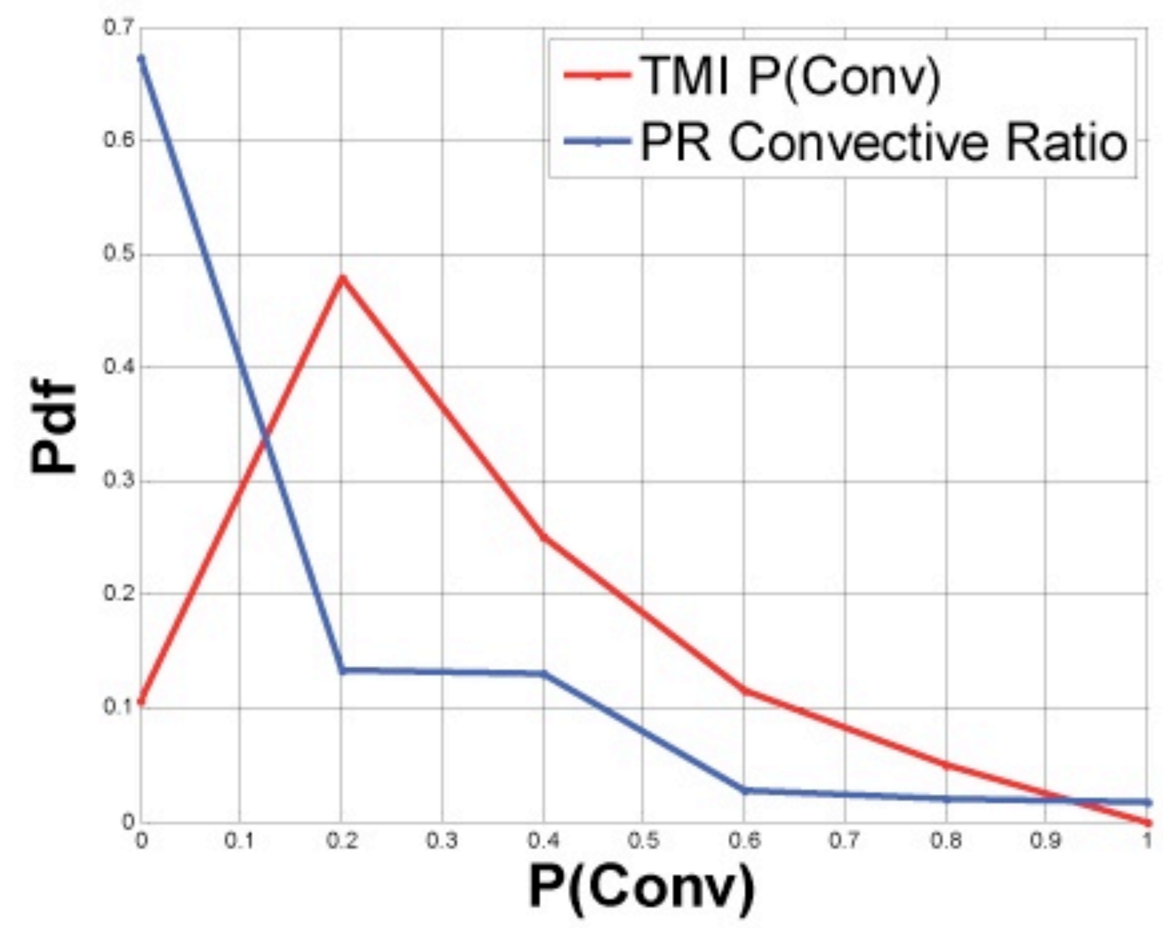
Radar

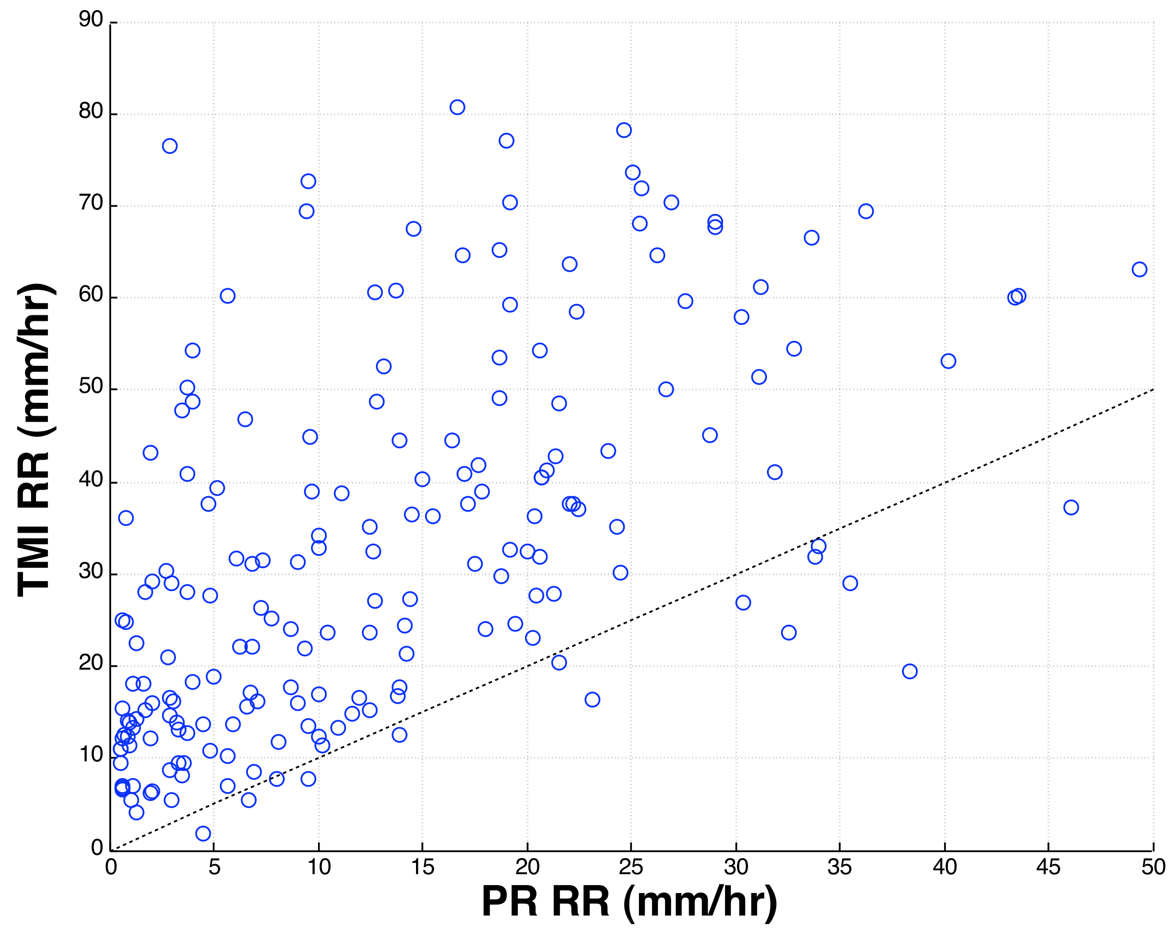
Microwave T_b 85_v GHz

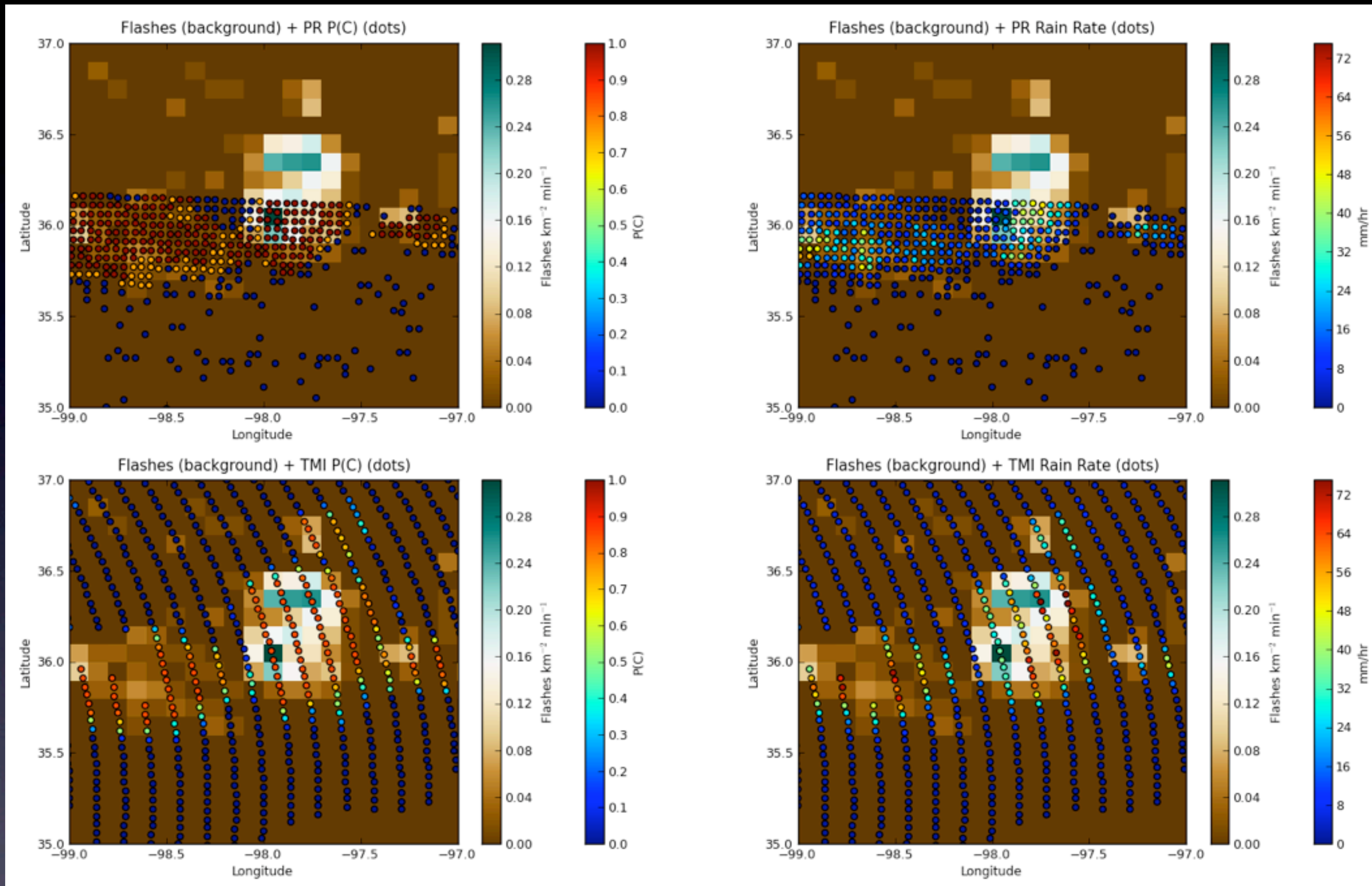
Flash centroid density matches cellular features from PR and TMI, but also emphasizes some cells over others, suggesting that lightning adds information content.

Gopalan et al. (2009) – analysis of over-prediction of rain rates by TMI

Too many small non-zero convective percentage classifcations (top) leads to overprediction of rain rates (bottom) for official TMI v6 algorithm







TMI-determined region of high convective percentage too extensive. Lightning more closely matches PR.

Other possible investigations

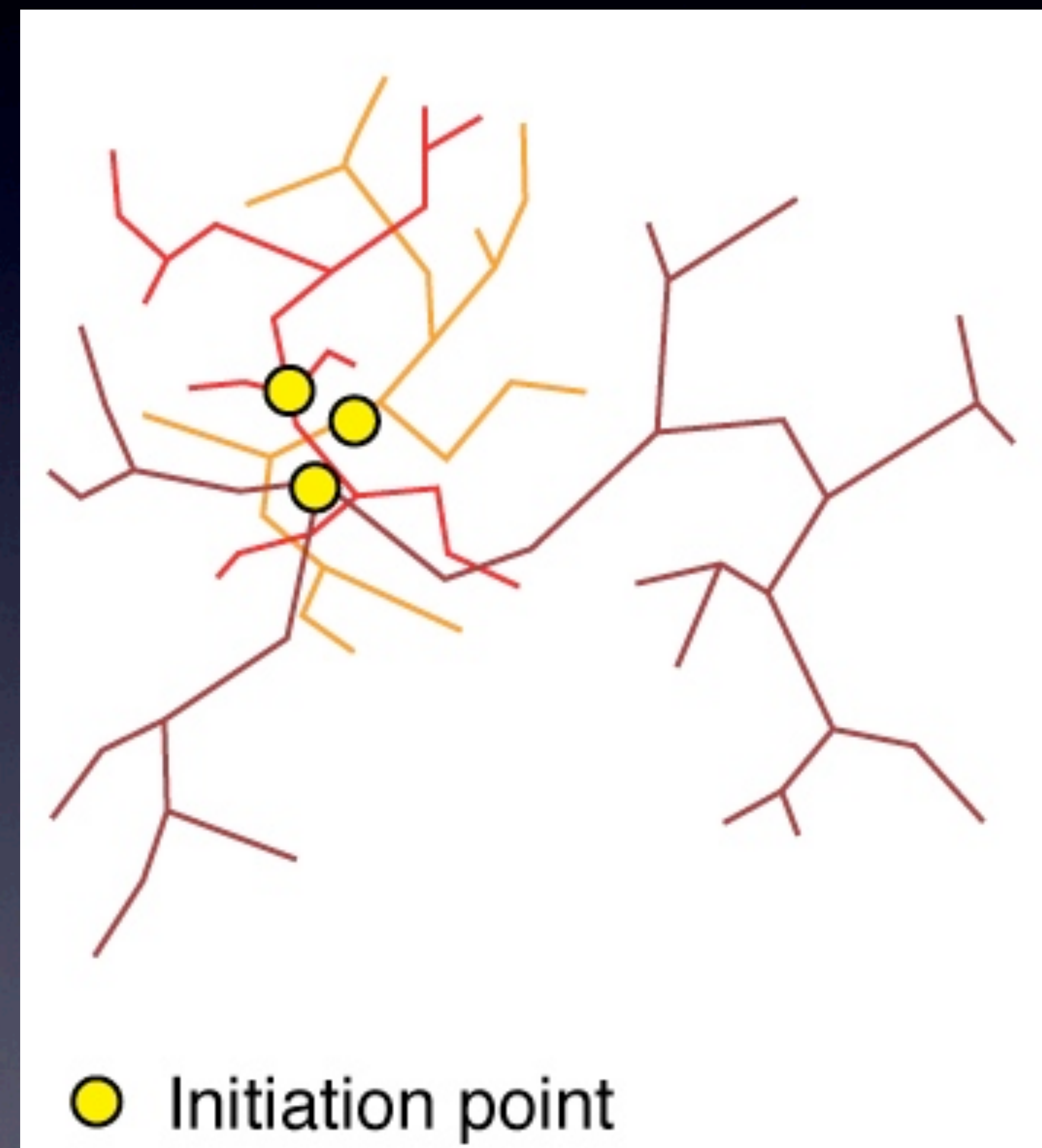
- Scatter plot comparisons of TMI, PR, and LIS parameters, to facilitate understanding of interrelationships that contribute to improved algorithm performance
- Detailed examination of case studies
- Operationalized version of algorithm improvements

Interpretation of lightning density patterns

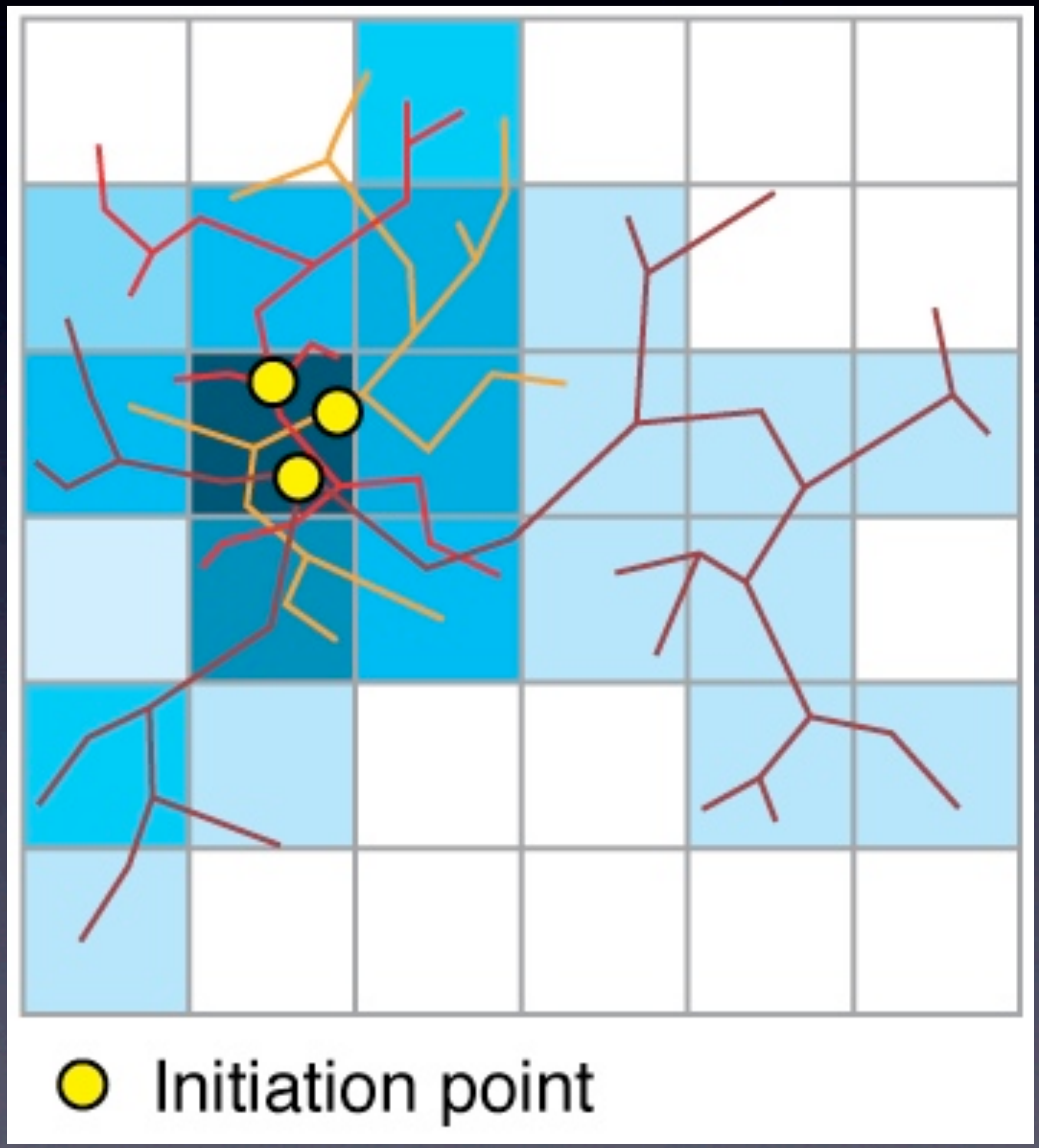
Interpretation of Total Lightning Density Patterns to Infer Storm Processes

- Total: all lightning in the cloud, including flashes that come to ground
    - On average, 5x more activity in cloud
- Density: column total detections
    - might be flashes, channel segments, optical pulses, etc.
    - patterns may differ depending on phenomenon detected, and are due to multiple physical processes

Picturing total lightning density

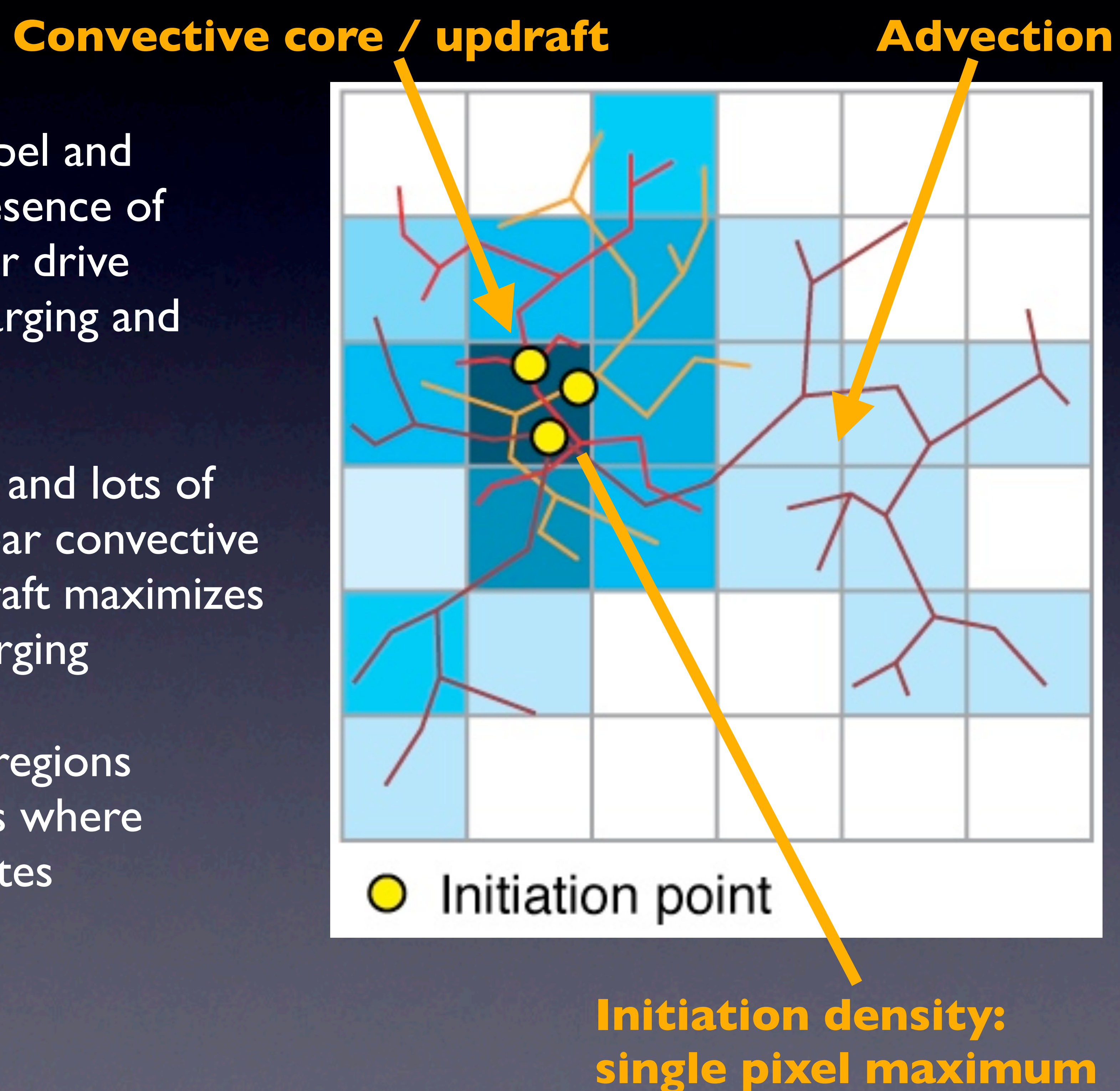


Picturing total lightning density



Picturing total lightning density

- Collisions of graupel and crystals in the presence of supercooled water drive thunderstorm charging and lightning.
- Maximum density and lots of flash initiations near convective cores where updraft maximizes microphysical charging
- Lower density in regions surrounding cores where advection dominates



Ongoing work

for end-user training in the GOES-R Proving Ground

Use a theoretical framework accessible to meteorologists to attribute lightning initiation and propagation to physical processes.

- Lightning **initiation** is most frequent when the local time tendency of the spatial derivative of charge density is maximized (large E-field), i.e.,
  - $|\partial(\nabla\rho)/\partial t| >> 0$ **Convective cores**
- Lightning will **propagate** more frequently through regions where relatively large values of charge density are maintained. Net charge density at some moment is the integral over all time of sources and sinks of net charge,
  - $\int(\partial\rho\ /\partial t)\ dt$ **Regions dominated by advection**
- Charge conservation for a single hydrometeor species is given by (Ziegler et al. 1991, Mansell et al. 2005):
  - $\partial\rho\ /\partial t = -\nabla\bullet(v\rho) - \partial(\rho V_t)/\partial z + \nabla\bullet(Kd\nabla\rho) + Sp$
  - – Flux Convg – Diff'l Sedimentation + Diffusion + Charging - Lightning

Next steps

- Would like some proxy FGE data to work with so that I can call it a true GLM proxy
- Posters at GOES-Users, AGU